

سیستم توصیه گر گردشگری مبتنی بر اعتماد با استفاده از خوشه‌بندی زمینه-آگاه

علی اکبر نوریان اول*، رضا روانمهر**، علی هارون آبادی***، فاطمه نوری****

تاریخ دریافت: ۱۳۹۸/۱/۱۹- تاریخ پذیرش: ۱۳۹۸/۷/۲۱

چکیده

امروزه در عصر ارتباطات تولید و انتقال اطلاعات با سرعتی روزافزون صورت می‌گیرد. در حوزه گردشگری نیز مقاصد زیادی وجود دارند که ارزش سفر کردن را دارند، اما افزایش تعداد مکان‌های مناسب گردشگری از میان حجم بالای انتخاب‌ها، انتخاب گزینه مطلوب را سخت می‌سازد؛ بنابراین اولویت‌بندی و شخصی‌سازی اطلاعات برای ارائه توصیه مناسب و مؤثر به صورت یک نیاز درآمده است. در این مقاله با در نظر گرفتن موقعیت مکانی کاربر هدف، یک سیستم توصیه گر گردشگری جهت پیشنهاد مکان مناسب برای اقامت گردشگر ارائه می‌گردد که از ترکیب اطلاعات زمینه، معیار اعتماد و خوشه‌بندی گراف، در جهت افزایش دقت پیشنهادها به کاربران استفاده می‌کند. ارزیابی روش پیشنهادی و مقایسه عملکرد آن با سایر روش‌های مورد آزمایش بر روی مجموعه داده Trip Advisor بیانگر آن است که این روش نتایج بهتری نسبت به سایر روش‌های مورد مقایسه نشان می‌دهد.

کلیدواژگان: سیستم توصیه گر گردشگری، خوشه‌بندی گراف، زمینه-آگاه، اعتماد، Advisor Trip

* . دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران

** . استادیار، گروه مهندسی کامپیوتر، واحد تهران مرکزی، دانشگاه آزاد اسلامی، ایران (نویسنده مسئول)

r.ravanmehr@iauctb.ac.ir

*** . استادیار، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران

**** . کارشناس ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران

مقدمه

طبق گزارش سال ۲۰۱۷ مجمع جهانی اقتصاد با ورود ۵,۲ میلیون گردشگر بین‌المللی به ایران و حدود ۳,۵ میلیارد دلار هزینه پرداخت‌شده توسط آن‌ها متوسط درآمد ایران به ازای هر گردشگر ۶۵۵ دلار بوده است. به‌طور کلی، صنعت گردشگری با ارزش‌افزوده حدود ۱۰,۱ میلیارد دلاری سهم ۲,۵ درصدی از کل تولید ناخالص داخلی ایران را تشکیل می‌دهد که با ۴۷۶ هزار شغل سهم ۱,۹ درصدی از کل اشتغال را در ایران به خود اختصاص داده است (سازمان جهانی گردشگری و شورای جهانی سفر و گردشگری^۱، ۲۰۱۷). مزیت رقابتی که در مقصدهای مختلف گردشگری در ایران وجود دارد می‌تواند این صنعت را به یکی از پربازده‌ترین صنایع کشور تبدیل کند (فرزین و نادعلی پور، ۱۳۹۰).

بررسی این آمار و ارقام اهمیت صنعت گردشگری را نشان می‌دهد برای توسعه این صنعت تنها داشتن جاذبه‌های گردشگری کافی نیست بلکه باید از الزامات فناوری در دنیای مدرن بهره جست. گردشگری الکترونیکی به کارگیری فن‌آوری‌های نوین اطلاعاتی و ارتباطی برای ارائه سرویس‌های موردنیاز به گردشگران است. بدون شک با فراهم شدن این فناوری‌ها در فضاهایی مانند شهر الکترونیک ارائه سرویس‌های موردنیاز به گردشگران ساده‌تر، باکیفیت‌تر و با هزینه کمتری انجام می‌شود. دستگاه‌های موبایل و اینترنت به گردشگران فرصت‌هایی عالی ارائه می‌دهند که به اطلاعات گردشگری دسترسی داشته باشند، اما افزایش عظیم تعداد انتخاب‌های گردشگری انتخاب گزینه موردنظر را سخت می‌سازد.

سیستم‌های توصیه‌گر را می‌توان به‌عنوان برنامه‌هایی توصیف کرد که تلاش دارند آیت‌های مناسب را به کاربران خاص توصیه کنند و این کار از طریق پیش‌بینی علایق کاربر نسبت به آیت‌ها، بر اساس اطلاعات مربوط به آیت‌ها، کاربر و تراکنش‌های بین آیت‌ها و کاربران صورت می‌گیرد. لذا، سیستم‌های توصیه‌گر می‌توانند در زمان کاربران جهت جستجوی اطلاعات صرفه‌جویی کنند. یکی از بیشترین کاربردهای سیستم‌های توصیه‌گر نیز در صنعت گردشگری است (کزاز و همکاران^۲، ۲۰۱۸). روش‌های مختلف توصیه در سیستم‌های توصیه‌گر اعمال

1. World Tourism Organization (UNWTO) and World Travel and Tourism Council (WTTC)

2. Kzaz et al.

شده‌اند و این کار بر اساس میزان پیچیدگی و نیاز آیت‌های توصیه انجام شده است. در دنیای واقعی، نمونه چنین سیستم‌های کاربردی را می‌توان پیشنهاد کتاب، دیسک فشرده^۱ و سایر محصولات در سایت Amazon.com، سرویس پیشنهاد فیلم در MovieLens (ریچی و همکاران^۲، ۲۰۱۱)، سرویس مرور و پیشنهاد آیت‌های تجاری در سایت Epinion (جینینگ^۳، ۲۰۱۰)، سرویس پیشنهاد اخبار در پروژه VERSIFI (شپیتسن و همکاران^۴، ۲۰۰۸) و سرویس توصیه در خصوص سفر، مکان‌ها و فعالیت‌های کاربران در سایت TripAdvisor (اسچین و همکاران^۵، ۲۰۰۲) دانست. این موارد از نمونه‌های اولیه و کارآمد سیستم‌های توصیه گر در حوزه تجاری محسوب می‌شوند که توانسته‌اند با افزودن قابلیت پیشنهادده به سرورهای تجاری خود، میزان فروش و رضایت مشتریان خود را افزایش دهند.

سیستم‌های توصیه گر گردشگری با استفاده از مجموعه داده‌های مختلف و الگوریتم‌های هوشمند می‌توانند در سریع‌ترین زمان ممکن بهترین مکان‌ها را برای بازدید یا اقامت به گردشگران پیشنهاد نمایند. برای اینکه این توصیه‌ها دقت و کاربری لازم را برای گردشگر هدف داشته باشد می‌توان از اطلاعاتی که گردشگران قبلی برای مقاصد مختلف در سایت‌های گردشگری نظیر Tripadvisor ثبت می‌کنند، استفاده کرد. در واقع، این اطلاعات مانند این است که از تجربه صدها و هزاران گردشگر قبلی برای انتخاب بهترین‌ها استفاده نمود و در نتیجه وابستگی‌های انسانی گردشگری نظیر استفاده از راهنمای سفر اندک می‌شود (ضیایی و فرمانی، ۱۳۸۹).

از چالش‌های سیستم‌های توصیه گر موجود میزان دقت و کارایی این سیستم‌ها در تخمین رتبه‌های کاربران به آیت‌های مختلف است. هر چه میزان دقت یک سیستم توصیه گر در تخمین رتبه‌های کاربران بالاتر باشد، سیستم بهتر قادر خواهد بود که پیشنهاد‌های بهتری متناسب با علایق کاربران به آن‌ها ارائه نماید. بر این اساس یکی از مهم‌ترین مسائلی که در سیستم‌های

1. CD

2. Ricci et al.

3. Jinming

4. Shepitsen et al.

5. Schein et al.

توصیه گر بایستی مورد توجه قرار گیرد، مسئله انتخاب همسایگی برای کاربر هدف است. اگر یک روش بتواند همسایگی های خوبی را برای کاربر هدف تعیین کند، قادر خواهد بود رتبه های مورد نظر را با دقت بالایی تخمین بزند و همچنین درصد رتبه هایی که پیش بینی می شود نیز افزایش می یابد. از جمله روش هایی که می تواند به سیستم های توصیه گر در فرآیند انتخاب همسایگی کمک کند، روش های مبتنی بر خوشه بندی می باشند.

در این مقاله تلاش شده در جهت افزایش دقت سیستم های توصیه گر گردشگری اقدام شود و با در نظر گرفتن موقعیت کاربر (گردشگر) هدف، یک سیستم توصیه گر گردشگری جهت پیشنهاد مکان مناسب برای اقامت گردشگر ارائه گردد. استفاده ترکیبی از اطلاعات زمینه ای و خوشه بندی گراف اعتماد برای دسته بندی کاربران و آیتم ها از مزایای این تحقیق است. اطلاعات زمینه ای مورد نظر در این مقاله موقعیت مکانی گردشگر است که نیاز به پیدا کردن مکان مناسب در شهر مقصد گردشگری دارد. معیار اعتماد یک ماتریس مربعی و به اندازه تعداد کاربران سیستم است که در واقع نشان می دهد هر کاربر به چه کاربرانی اعتماد دارد. در واقع بیان کننده این مفهوم است که گردشگران قبلی که از مکان های مختلف گردشگری بازدید کرده و به آن ها امتیاز داده اند تا چه اندازه می توانند در انتخاب مکان مناسب برای گردشگران فعلی مورد استفاده واقع شوند. با گسترش روزافزون شبکه های اجتماعی و حضور مکان های گردشگری مقاصد مختلف در این شبکه ها، کاربران به سادگی می توانند به این رسانه ها مراجعه کرده و نظرات خود را در قالب متن، like و unlike ثبت کنند.

روش پیشنهادی از سه فاز کلی تشکیل شده است. ابتدا در فاز اول، پس از پیش پردازش و استخراج کاربران و آیتم های متفاوت، اطلاعات استخراج شده در یک پایگاه داده به منظور استفاده های بعدی ذخیره می گردد، سپس در فاز دوم، بر اساس گراف اعتماد ساخته شده از اطلاعات زمینه ای، خوشه بندی صورت می گیرد و در فاز سوم، با جمع آوری رتبه های مورد نیاز از کاربران مورد اعتماد، مقدار رتبه برای هر یک از هتل ها به دست می آید و در نهایت بر اساس رتبه های پیش بینی شده برای کاربر هدف، N آیتم با رتبه بالاتر به عنوان هتل های مورد علاقه کاربر هدف به وی پیشنهاد می شوند. روش پیشنهادی به سادگی می تواند به برنامه ای کاربردی

تحت موبایل یا وب به عنوان یک ابزار کاربردی گردشگری استفاده شود (منصوری مؤید و سلیمانی، ۱۳۹۱).

ساختار مقاله از شش بخش تشکیل شده است. پس از مقدمه که در بخش اول آمده است، در بخش دوم، مفاهیم بنیادین سیستم‌های توصیه گر از جمله روش‌های تولید پیشنهاد و همچنین نقاط ضعف و قوت هر کدام بررسی می‌گردد. در بخش سوم کارهای مرتبط مورد مطالعه و بررسی قرار گرفته‌اند و روش‌ها، الگوریتم‌ها، متدها و چارچوب‌های معرفی شده مطرح می‌شوند. در بخش چهارم به معرفی رویکرد پیشنهادی مبتنی بر ارائه یک روش زمینه آگاه جهت افزایش دقت سیستم‌های توصیه گر گردشگری می‌پردازد. در بخش پنجم، مجموعه‌ای از آزمایش‌های مختلف به منظور ارزیابی کارایی روش پیشنهادی و مقایسه آن با سایر روش‌های مورد ارزیابی بیان خواهد شد. بخش ششم به نتیجه‌گیری کلی مقاله و ارائه پیشنهادهایی جهت توسعه و بهبود روش‌های پیشنهادی برای کارهای آینده می‌پردازد.

مبانی نظری

سیستم‌های توصیه گر

سیستم‌های توصیه گر به عنوان نمونه‌ای از سیستم‌های تصمیم‌یار^۱ شناخته می‌شوند که در دهه ۹۰ میلادی به عنوان شاخه‌ای مستقل پا به عرصه تحقیق و پژوهش گشودند. از این سیستم‌ها به عنوان یکی از راه‌حل‌های رویارویی با مسئله سرباری اطلاعات در تجارت الکترونیک یاد می‌شود (اسچلیت و همکاران^۲، ۱۹۹۴). در یک سیستم توصیه گر با n کاربر و m آیت، مجموعه کاربران را به صورت $U = [u_1, u_2, \dots, u_n]$ و مجموعه آیت‌ها را با $I = [i_1, i_2, \dots, i_m]$ نشان می‌دهم. ماتریس کاربر-آیت از ورودی‌های اصلی یک سیستم توصیه گر است. این ماتریس شامل رتبه کاربران به آیت‌های موجود در سیستم است. به عبارت دیگر، هر سطر نشان‌دهنده یک کاربر و هر ستون یک آیت است. ممکن است کاربران به تعداد کمی از آیت‌ها رتبه داده باشند. هر عنصر در فضای مجموعه U ، می‌تواند توسط پروفایل کاربران با

1. Decision Support System

2. Schilit

مشخصه‌های متفاوتی از قبیل سن، جنسیت، میزان درآمد، وضعیت تأهل و غیره مشخص شود. در ساده‌ترین حالت، پروفایل کاربران فقط شامل یک عنصر منحصر به فرد به نام شناسه کاربر است. به‌طور مشابه در فضای مجموعه I، هر آیتم می‌تواند توسط مشخصه‌های آن تعریف گردد.

از جمله تعاریفی که برای سیستم‌های پیشنهاددهنده شده به این صورت است که:

- زیرمجموعه‌ای از سیستم‌های تصمیم‌یار و نوعی سیستم اطلاعاتی که توانایی تحلیل رفتارهای گذشته و ارائه توصیه‌هایی برای مسائل جاری را دارا هستند.
- شیوه تفکر کاربر را به کمک اطلاعاتی که از نحوه رفتار او یا کاربران مشابهش در اختیار دارد حدس می‌زند و مناسب‌ترین آیتم به سلیقه‌اش را شناسایی و پیشنهاد می‌کند.

از یک دیدگاه خاص سیستم‌های توصیه‌گر را می‌توان به دو دسته برخط و برون‌خط^۱ تقسیم‌بندی کرد. در فاز برون‌خط، سیستم، اطلاعات مربوط به کاربران را پردازش و تجزیه و تحلیل کرده و مدلی از علایق هر کاربر را تشکیل می‌دهد. سپس در فاز برخط با استفاده از این مدل، علایق کاربر را تشخیص داده و با ارائه پیشنهادهای مناسب، از اتلاف وقت او بر روی وب جلوگیری می‌کند. از طرفی دیگر مدل‌های پایه در سیستم‌های توصیه‌گر بر اساس دو دسته از داده‌ها، تعریف شده و مورد استفاده قرار می‌گیرند (کلهر و بریج^۲، ۲۰۰۴): بر اساس تعاملات کاربر و آیتم، مانند رتبه‌ها یا رفتار خرید و اطلاعات در مورد کاربران و آیتم‌ها، مانند کلمات کلیدی مرتبط یا نمایه متنی^۳. روش‌هایی که از تعاملات کاربران و آیتم‌ها استفاده می‌کنند به‌عنوان روش‌های پالایش گروه شناخته می‌شوند؛ درحالی‌که روش‌هایی که از مدل دوم استفاده می‌کنند به‌عنوان روش‌های مبتنی بر محتوا شناخته می‌شوند (لوکاس و همکاران^۴، ۲۰۱۳).

1. Offline

2. Kelleher & Bridge

3. Textual Profile

4. Lucas et al.

سیستم‌های پیشنهاددهنده باید پیشنهادهای خود را با در نظر گرفتن موارد زیر به کاربر ارائه دهند:

نیاز کاربر و دانش سیستم در مورد وی

شرایط و محیطی که کاربر در آن قرار دارد.^۱

یکی از مهم‌ترین روش‌های مورد استفاده در سیستم‌های توصیه گر، روش پالایش گروه^۲ است (یوچگیت و کلارک^۳، ۲۰۰۴). این روش به منظور تولید پیشنهادهای مناسب برای کاربر هدف، از نظرات کاربران مشابه با کاربر هدف (به‌عنوان کاربران همسایه) در فرآیند تولید پیشنهاد استفاده می‌کند. کاربر هدف، کاربری است که سیستم، پیشنهادهایی را به او ارائه می‌دهد. برای این منظور، روش پالایش گروه از رتبه‌هایی که قبلاً توسط کاربر هدف به آیتم‌های مختلف داده شده است، برای محاسبه شباهت بین کاربران استفاده می‌کند. معیار شباهت^۴، میزان شباهت بین دو کاربر را مشخص می‌کند و بر اساس آن همسایگان کاربر هدف تعیین می‌شوند. روش پالایش گروه به دو دسته مبتنی بر حافظه^۵ و مبتنی بر مدل^۶ تقسیم می‌شود (وانگ و همکاران^۷، ۲۰۱۰).

سیستم‌های توصیه گر زمینه آگاه^۸

اخیراً در برخی سیستم‌های توصیه گر اطلاعات زمینه‌ای در نظر گرفته شده است. از اطلاعات زمینه‌ای مانند زمان، مکان، یا افرادی که کاربر هم‌اکنون با آن‌ها (دوستان، خانواده یا همکاران) همراه است می‌توان اشاره کرد (هارونا و همکاران^۹، ۲۰۱۷).

1. User Context
2. Collaborative Filtering
3. Uchyigit & Clark
4. Similarity Measure
5. Memory-Based Methods
6. Model-Based Methods
7. Wang
8. Context Awareness-Based
9. Haruna

اطلاعات زمینه‌ای اطلاعاتی کامل‌تری برای ایجاد توصیه فراهم می‌آورند به‌ویژه برای برخی برنامه‌های کاربردی که در آن‌ها، در نظر گرفتن کاربر و آیتم‌ها به‌تنهایی کافی نیست، مانند توصیه کردن یک پکیج تعطیلات، یا محتوای شخصی‌سازی شده در یک وب‌سایت. در اینجا بسیار اهمیت دارد که اطلاعات متنی را در پروسه توصیه بگنجانیم تا در شرایط خاص بتواند آیتم‌های مناسبی را به کاربر توصیه کند. به‌عنوان مثال، با استفاده از زمینه زمانی، سیستم توصیه‌کننده، سفر بسیار متفاوتی در زمستان نسبت به تابستان به کاربر پیشنهاد می‌دهد. هدف سیستم‌های پیشنهاددهنده زمینه آگاه ارائه سیستم‌های پیشنهاددهنده با کیفیت بهتری به کمک مشارکت دادن اطلاعات در دسترس زمینه‌ای کاربر است. در حقیقت زمینه به‌عنوان یک بعد به سیستم افزوده می‌شود (ویکتراس و همکاران، ۲۰۱۸).

با وجود پیشرفت‌های که در فناوری حسگرها پدید آمده است، ترکیب آن‌ها با فناوری‌ها و وسایل سیار، برای کاربردهای آگاه از زمینه بسیار سودمند است. از این فناوری به‌طور گسترده در علوم چون رباتیک و بینایی ماشین استفاده شده و با کاهش هزینه و اندازه این حسگرها، امکان به‌کارگیری آن‌ها در وسایل سیار، بیشتر از قبل ممکن شده است. با وجود انواع زیاد آن‌ها، پارامترهای زمینه‌ای زیادی چون دمای هوا، فشار هوا، میزان رطوبت، شدت نور، طول موج، تشخیص حرکت و لرزش، پارامترهای مرتبط با مکان مثل افراد و اشیاء واقع در یک محدوده مکانی، جهت، سرعت و پارامترهای فردی چون فعالیت‌های روزانه یا مسیر تردد که با کمک آن‌ها می‌توان به وضعیت و شرایط او پی برد، قابل در نظر گرفته شدن در سیستم‌های پیشنهاددهنده می‌باشند.

فرایند توصیه به‌طور معمول با تعیین مجموعه اولیه رأی‌ها که به‌طور صریح یا به‌طور ضمنی از سیستم جمع‌آوری شده، آغاز می‌شود. هنگامی که این رأی‌های اولیه مشخص شدند، سیستم پیشنهاددهنده شروع به تخمین تابع رأی‌دهنده برای جفت (کاربر، آیتم) ای که هنوز توسط کاربر رأی نگرفته است، می‌کند. زمانی که تابع برای کل فضای تخمین زده شد، سیستم پیشنهاددهنده می‌تواند آیتم‌های با بیشترین مقدار رأی را به هر کاربر پیشنهاد کند. چنین سیستم‌های سنتی یا دوبعدی نامیده می‌شود زیرا آن‌ها دو بعد کاربر و آیتم را در فرآیند

توصیه در نظر می گیرند. به عبارت دیگر، در متداول ترین فرم مدل سازی، مسئله توصیه به مسئله تخمین رأی، برای آیتم های که توسط کاربر دیده نشده اند، تقلیل می یابد. این تخمین ها به طور معمول مبتنی بر رأی های همین کاربر به سایر آیتم ها (روش های مبتنی بر محتوا) یا رأی-های داده شده توسط سایر کاربرها به این آیتم ها (روش های پالایش گروهی) است. با وجود اینکه مطالعه های زیادی در زمینه سیستم های پیشنهاددهنده انجام گرفته است، اغلب رویکردهای موجود این سیستم ها، اطلاعات زمینه ای از جمله مکان، زمان و... را در نظر نمی گیرند ولی در سیستم های پیشنهاددهنده زمینه آگاه، از اطلاعات زمینه ای در فرآیند توصیه استفاده می شود. در این حالت رأی های کاربرها، به عنوان تابعی از کاربرها، آیتم ها و زمینه مدل سازی می شوند.

مشخص است که زمینه های کاربر بر ترجیحات او اثر می گذارند. هدف سیستم های پیشنهاددهنده زمینه آگاه ارائه سیستم های پیشنهاددهنده با کیفیت بهتری به کمک مشارکت دادن اطلاعات در دسترس زمینه ای کاربر است. در حقیقت زمینه به عنوان یک بعد به سیستم افزوده می شود. در ابتدا سیستم های پیشنهاددهنده زمینه آگاه از فاکتور مکان به عنوان زمینه استفاده می کردند بعدها سایر زمینه ها مانند تاریخ، فصل، دما و وضعیت روحی نیز به عنوان زمینه مورد بررسی قرار گرفت. فقط در بحث مکان، دسته مهمی از سیستم های زمینه آگاه را سیستم-های آگاه از مکان تشکیل می دهند. کم کم استفاده از سایر اطلاعات زمینه ای علاوه بر مکان نیز مورد توجه توسعه دهندگان این نوع سیستم ها قرار گرفته است. به طور مثال یکی از حوزه های جذاب برای پیاده سازی سیستم های پیشنهاددهنده زمینه آگاه، گردشگری است. امروزه گردشگران انتظار دارند که دسترسی شخصی به اطلاعات گردشگری در هر زمان، مکان و هر شرایطی را داشته باشند. سیستم های پیشنهاددهنده گردشگری سیار، چنین اطلاعاتی را در اختیار کاربران قرار می دهند.

سیستم های توصیه گر مبتنی بر خوشه بندی

یکی از مهم ترین مشکلات و چالش های موجود در سیستم های توصیه گر، انتخاب همسایگی برای کاربر هدف به منظور پیش بینی رتبه های این کاربر است. سیستم های توصیه گر پالایش گروه که بر اساس کاربران همسایه کاربر هدف، رتبه ها را پیش بینی می کنند، باید قادر باشند که

کاربران همسایه کاربر هدف را به خوبی شناسایی نمایند. در این گونه سیستم‌ها، با انتخاب درست همسایگی کاربران، می‌توان برای کاربر هدف رتبه‌هایی با دقت بالا را پیش‌بینی نمود. انتخاب و تعیین کاربران همسایه در سیستم‌های توصیه‌گر با چالش‌هایی روبه‌رو است. یکی از مشکلاتی که در این زمینه وجود دارد، مشکل تنگی داده^۱ و کاربران شروع سرد^۲ است. در صورتی که یک کاربر هدف به تعداد کمی از آیتم‌ها رتبه داده باشد، محاسبه شباهت بین این کاربر و سایر کاربران موجود در سیستم با مشکل روبه‌رو خواهد شد. در نتیجه، انتخاب همسایگی مناسب برای کاربر هدف کار راحتی نخواهد بود. یکی از روش‌هایی که به منظور رفع مشکل انتخاب همسایگی در سیستم‌های توصیه‌گر مورد استفاده قرار گرفته است، روش خوشه‌بندی است (هراندو و همکاران^۳، ۲۰۰۷). خوشه‌بندی باعث می‌شود که کاربران و یا آیتم‌های مشابه در یک خوشه قرار بگیرند؛ بنابراین، می‌توان از کاربران موجود در خوشه کاربر هدف برای پیش‌بینی رتبه مورد نظر استفاده کرد.

پیشینه پژوهش

حجم اطلاعات مربوط به سیستم‌های گردشگری و جهانگردی در اینترنت روزبه‌روز در حال افزایش است. سیستم‌های توصیه‌گر گردشگری می‌توانند اطلاعات را برای هر فرد، شخصی-سازی کنند. برخی از این سیستم‌ها بر جاذبه‌ها و مقصدها تمرکز دارند، درحالی‌که سایرین برنامه‌های سفر که شامل حمل‌ونقل، رستوران‌ها و محل اقامت است، پیشنهاد می‌دهند. روش‌های مختلف توصیه‌ای در سیستم‌های توصیه‌گر گردشگری اعمال شده‌اند و این کار بر اساس میزان پیچیدگی و نیاز آیتم‌های توصیه‌شده می‌باشند.

در مقاله (لیال^۴ و همکاران، ۲۰۱۷) مدل‌سازی مبتنی بر اعتماد از رتبه‌بندی چندمعیاره برای ایجاد توصیه‌های آنلاین مبتنی بر کاربر با استفاده از پالایش گروه انجام گرفته است. نمایه‌سازی، رتبه‌بندی‌های چندمعیاره را به منظور به دست آوردن بهترین و پالایش‌شده‌ترین پروفایل بررسی

1. Data Sparsity

2. Cold Start Users

3. Hernando

4. Leal

می‌نماید. در این مقاله بهترین نمونه‌های رتبه‌بندی که معنی‌دارترین رتبه‌بندی‌های چندگانه است بعد از تحلیل همبستگی^۱ برای شناسایی رابطه بین رتبه‌بندی‌های مختلف با استفاده از رگرسیون خطی چندگانه^۲ تحلیل می‌شوند. بر طبق مجموعه داده استفاده شده در این تحقیق، کاربران مهمان‌های هتل و آیت‌ها، هتل‌ها هستند.

سیستم‌های توصیه گر گردشگری تحت موبایل در حال حاضر کاربرد زیادی دارند چون اطلاعات زمینه‌ای مانند موقعیت مکانی و زمانی در اختیار سیستم‌ها می‌گذارند. در مقاله (گاوالاس و همکاران^۳، ۲۰۱۵) یک سیستم توصیه گر زمینه آگاه بنام eCOMPASS ارائه شده است. این سیستم یک برنامه کاربردی زمینه آگاه موبایل برای ارائه برنامه سفر روزانه است که به گردشگران توصیه‌های مربوط به استفاده از وسایل نقلیه عمومی با در نظر گرفتن تأخیرهای وارده در ایستگاه‌های حمل‌ونقل ارائه می‌دهد. زمینه مدنظر در این مقاله زمان و مکان بوده که سیستم از این اطلاعات همراه با اخذ ترجیحات صریح کاربر برای زمان در اختیار بهره می‌برد.

در پژوهش (تیواری و کواشیک^۴، ۲۰۱۴) غنی‌سازی مکان‌های گردشگری با استفاده از رویکرد سنجش جمعیت که توسط سیستم توصیه گر نقاط گردشگری برای کاربران تلفن همراه استفاده می‌شود، بررسی شده است. در این پژوهش توصیه‌های زمینه آگاه و نقاط جذاب گردشگری برای کاربران تلفن همراه با توجه به نزدیکی مکان جاری فراهم می‌کند.

در مقاله (هوانگ^۵، ۲۰۱۶) بررسی روش‌های زمینه آگاه برای تهیه یک سیستم پیشنهاددهنده مکانی با توجه به زمینه‌های مثل آب‌وهوا، فصل، ساعت و روز برای مراکز گردشگری شبکه‌های اجتماعی است. در اینجا تاریخچه سفر هر کاربر توسط سیستم ایجاد می‌شود. پژوهشگر دو جنبه مهم را لحاظ کرده است: یک اینکه بجای انتخاب زمینه‌های دقیقاً مشابه کاربر جاری مجموعه انتخاب‌های کاربران مشابه را بر اساس معیار شباهت پالایش گروهی بین زمینه‌ها برمی-

1. Correlation Analysis

2. Multiple Linear Regression

3. Gavalas

4. Tiwari & Kaushik

5. Huang

گزینند. دوم اینکه پس از انتخاب زمینه‌های مشابه بر اساس پالایش گروهی، نسبت به پیشنهاد معیار شباهت جدیدی بین زمینه‌ها اقدام کرده و سیستم را بر اساس سه روش پیشنهاددهنده زمینه آگاه یعنی پیش پایش زمینه‌ای، پس پایش زمینه‌ای و مدل‌سازی زمینه می‌سازد. یکی از ضعف‌های سیستم‌های توصیه‌گر گردشگری مبتنی بر پالایش گروه این است که معمولاً این سیستم‌ها فقط از اطلاعات رتبه‌های پیشین کاربران استفاده نموده و از یک جنبه تلاش می‌کنند سیستم را پیش‌بینی نمایند. برای حل این مشکل در مقاله (آزادجلال و همکاران^۱، ۲۰۱۷) تلاش شده است تا با در نظر گرفتن جنبه‌های مختلف و اطلاعات اجتماعی و آنلاین کاربران دقت پیش‌بینی را بهبود بخشد. در این مقاله یک سیستم توصیه‌گر جهت توسعه سیستم‌های پالایش گروه چندمعیاره^۲ جهت پیشنهاد هتل ارائه شده است و برای بهبود پیش‌بینی از مدل ترکیبی گوسی^۳ با ترکیب الگوریتم **بیشینه‌سازی انتظار**^۴ و سیستم استنتاج عصبی-فازی تطبیقی^۵ استفاده شده است تا بر اساس این دو الگوریتم رتبه‌های نهایی با دقت بالاتری پیش‌بینی شود. برای کاهش ابعاد نیز از تجزیه و تحلیل مؤلفه اصلی^۶ برای پاسخگویی به چندین خطای ناشی از وابستگی متقابل بین معیارها در مجموعه داده‌های چند معیاره استفاده شده است. هدف اصلی در پژوهش (محمدنژاد و مهدوی، ۲۰۱۲) تولید پیشنهادها با کیفیت بالا است حتی زمانی که داده‌های در دسترس کافی نباشند. مدل ارائه شده در این مقاله از ترکیب روش‌های توصیه (به‌عنوان مثال ترکیب روش‌های توصیه پالایش گروه و مبتنی بر محتوا) با روش‌های دیگر مانند خوشه‌بندی و قوانین انجمنی استفاده کرده است. مدل مربوطه شامل چهار مرحله است، در مرحله اول گردشگران بر اساس مکان خود خوشه‌بندی می‌شوند و خوشه گردشگر هدف به مرحله بعدی فرستاده می‌شود. در مرحله دوم یک گراف دوسطحی بر اساس شباهت بین علایق گردشگر و شباهت مسیرها ساخته می‌شود. با توجه به این گراف، رابطه‌های تعدی

1. Azadjalal

2. Multi-criteria

3. Gaussian Mixture Model(GMM)

4. Expectation Maximization(EM)

5. Adaptive Neuro-Fuzzy Inference System (ANFIS)

6. Principal Components Analysis (PCA)

بین گردشگران و تعداد k از آیتم‌ها که دارای بالاترین وزن از روابط می‌باشند به گردشگران هدف پیشنهاد می‌گردد. با توجه به آزمایش‌ها، آزمون استاندارد اندازه‌گیری F نشان داد که کیفیت توصیه‌ها در این روش از روش‌های سنتی که قابلیت کشف روابط تعدی را ندارند بالاتر است.

در مقاله ژندینگ و همکاران^۱ (۲۰۱۵)، مکان‌های گردشگران با استفاده از نزدیکی فضایی عکس‌های برچسب‌گذاری شده جغرافیایی کشف می‌شوند. مکان‌های گردشگری که عکاسی شده‌اند به صورت یک مسئله خوشه‌بندی تلقی شده است و از الگوریتم‌های مانند k -means با استفاده از برچسب‌های عکس‌های مکان‌ها در این مقاله بهره برده شده است. پروفایل مکان‌ها برای توصیف اطلاعات زمینه‌ای فصل و آب‌وهوا ایجاد شده و برای پیدا کردن کاربران با اولویت‌های مشابه، تشابه کاربر-کاربر میان کاربران بر اساس توزیع موضوع تاریخچه‌های سطح کاربر محاسبه می‌شود. برای ایجاد توصیه برای یک کاربر در شهر مورد هدف نیز، ابتدا N کاربری که بیش از همه به هم شباهت دارند را پیدا می‌کند و سپس مکان‌های گردشگری را با استفاده از فیلترینگ مشارکتی پیدا کرده و این مکان‌ها بر اساس محدودیت‌های زمینه فیلتر می‌شود. در نهایت m مکان برتر را به عنوان نتیجه جستجوی کاربر هدف در شهر هدف برمی‌گرداند.

در پژوهش کاظمی و همکاران (۱۳۸۴)، مدلی مفهومی ارائه شده است که نشان می‌دهد چگونه می‌توان از سیستم‌های تصمیم‌یار در بخش اسکان صنعت گردشگری، جهت ارائه خدمات بهتر به گردشگران استفاده نمود. این سیستم از بخش‌هایی همچون پایگاه مدیریت مدل، پایگاه مدیریت داده و رابط - کاربر تشکیل شده که کاربر می‌تواند به راحتی شرایط خود را از طریق رابط - کاربر به سیستم منتقل کند و با استفاده از مدل، از بین گزینه‌های موجود در پایگاه داده‌های خود، با توجه به شرایط کاربر مناسب‌ترین گزینه‌ها که رضایت و مطلوبیت بیشتری را برای او فراهم می‌نماید به وی، پیشنهاد کند. گردشگر می‌تواند از میان گزینه‌های پیشنهادشده، موردی را که بیشتر برایش جذاب بوده، انتخاب نماید و به راحتی امکان دسترسی به مشخصات کامل و ذخیره اتاق در اقامتگاه مطلوب خود را خواهد داشت.

در مقاله سهرابی و همکاران (۱۳۹۲)، به چگونگی طراحی و استفاده از سیستم خبره فازی برای انتخاب هتل محل اقامت پرداخته شده است. از هفت مرحله تعیین معیارهای تصمیم‌گیری، طراحی مجموعه‌های فازی معیارها، تبیین قواعد سیستم خبره، فازی سازی ورودی‌ها، استنتاج فازی، فازی زدایی و رتبه‌بندی گزینه‌ها تشکیل شده است. استفاده از منطق فازی، این امکان را می‌دهد که چنانچه یک هتل در یک زیرمعیار بدون انطباق لازم با شرایط مطلوب گردشگر بود، وضعیت زیرمعیار متناظر دیگری بررسی شده تا بخت انتخاب هتل مناسب به آسانی از دست نرود. از سوی دیگر، قواعد سیستم خبره فازی، شرایط هتل‌ها را با توجه به وضعیت مطلوب گردشگر بررسی می‌کند و بدین ترتیب است که رتبه‌بندی هتل‌ها بدون مقایسه مستقیم گزینه‌ها، بلکه با مقایسه هر هتل با شرایط مطلوب گردشگر انجام می‌گیرد. وجود چنین سیستمی علاوه بر کمک به گردشگران و افزایش رضایت‌مندی آنان، به توسعه صنعت هتلداری تهران نیز کمک خواهد کرد.

در تحقیق بهاری سجه‌رود و طالعی (۱۳۹۴)، گردشگری به‌عنوان یکی از منابع اقتصادی، به‌خصوص در کشورهای دارای سابقه فرهنگی و جاذبه‌های متعدد گردشگری مورد تأکید قرار گرفته و یک سامانه برنامه‌ریزی گردشگری مبتنی بر وب جهت کمک به گردشگران برای بازدید از جاذبه‌های موردعلاقه در حداقل زمان ممکن، طراحی و پیاده‌سازی شده است. در این سامانه پیشنهادی از تلفیق روش‌های مبتنی بر سامانه حامی تصمیم‌گیری مکانی و توابع تحلیل مکانی استفاده شده است. سامانه طراحی شده از طریق اطلاعاتی از قبیل علائق، تعداد روزها و مکان شروع حرکت گردشگر، برای هرروز به‌صورت جداگانه، برنامه‌ریزی گردشگری را انجام داده و همراه با ارائه طرح روزانه گردشگر، بهترین مسیر بین مکان‌های منتخب را تعیین می‌نماید و مدیریت زمانی و مکانی به‌صورت هم‌زمان انجام می‌دهد.

در مقاله ابوالحسینی و آل شیخ (۱۳۹۷)، یک سامانه همراه طراحی، توسعه، اجرا و ارزیابی شده که امکان توصیه گردشگرهای مشابه به یکدیگر و برنامه‌ریزی گشت‌های تفریحی مشترک بر اساس موقعیت‌های جغرافیایی آن‌ها را میسر می‌کند. همراه بودن این سامانه تأثیر بسزایی در عملکرد آن می‌گذارد. این سامانه از یک سیستم توصیه‌گر مبتنی بر فیلتر جمعیت‌شناسی به‌منظور توصیه گردشگرهای مشابه به یکدیگر استفاده می‌کند. همچنین الگوریتم مورچه که

یکی از الگوریتم‌های مشهور هوش مصنوعی است به منظور برنامه‌ریزی مسیر در این سامانه مورد استفاده قرار گرفته است.

روش شناسی پژوهش

هدف اصلی روش پیشنهادی، استفاده از معیار اعتماد، اطلاعات زمینه و خوشه‌بندی گراف به منظور افزایش دقت پیش‌بینی رتبه‌ها در سیستم‌های توصیه گر گردشگری است. با توجه به شکل (۱)، روش پیشنهادی از سه فاز کلی تشکیل شده است.



شکل ۱. ساختار روش پیشنهادی به منظور پیش‌بینی رتبه‌ها

پیش پردازش داده‌ها

همان‌طور که در شکل (۱) دیده می‌شود در فاز اول به ساخت سه ماتریس رتبه، اعتماد و زمینه از مجموعه داده اصلی خواهیم پرداخت:

- **ماتریس رتبه‌ها:** این ماتریس بر اساس اطلاعات ثبت شده از رتبه‌های پیشین کاربران، ارائه شده است، در واقع ماتریس رتبه‌ها قسمتی از مجموعه داده اصلی است که بر اساس نوع مجموعه داده از سایت مربوطه استخراج می‌شود. در سیستم توصیه گر گردشگری TripAdvisor از کاربران خواسته می‌شود به مکان‌هایی که قبلاً دیده‌اند و موردپسند آن‌ها قرار گرفته‌اند امتیاز دهند، بنابراین استفاده از داده این سایت می‌تواند برای ارزیابی روش پیشنهادی مناسب باشد. وانگ و همکاران در مجموعه داده^۱ مربوط به TripAdvisor را که شامل ۱۲۷۷۳ هتل و ۱۶۲۱۵۲۷ نظر درباره هتل‌ها است فراهم کرده است. در روش پیشنهادی از نسخه خلاصه شده این مجموعه داده در قالب JSON استفاده و ماتریس رتبه استخراج شده از آن با ابعاد ۱۲۳*۶۴۸ را جهت ارزیابی استفاده نموده‌ایم.

- **ماتریس اعتماد:** ماتریس اعتماد یک ماتریس مربعی و به اندازه تعداد کاربران سیستم است که در واقع نشان می‌دهد هر کاربر به چه کاربرانی اعتماد دارد. اعتماد هر کاربر به کاربر دیگر بر اساس یک عدد مشخص می‌شود. در صورت وجود ارتباط، مقدار اعتماد برابر با ۱ و در غیر این صورت مقدار اعتماد برابر با ۰ خواهد بود. ماتریس اعتماد در برخی از مجموعه داده‌ها همراه با ماتریس رتبه‌ها ارائه می‌شود اما در بیشتر مجموعه داده‌ها مانند مجموعه داده مورد استفاده ما مقدار اعتماد کاربران در آن‌ها به‌طور صریح مشخص نگردیده است لذا با توجه به معیار شباهت پیرسون (رابطه ۱)، از میان تمامی کاربران، کاربرانی که مقدار شباهت آن‌ها با توجه به رابطه (۲) از یک آستانه^۲ از پیش تعیین شده بیشتر باشد، به عنوان کاربران مورد اعتماد در نظر گرفته خواهند شد.

$$S_{\text{Pearson}}(a, u) = \frac{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^n (r_{u,i} - \bar{r}_u)^2}} \quad (1)$$

1. <http://times.cs.uiuc.edu/~wang296/Data/>

2. Threshold

که در این رابطه $S_{\text{Pearson}}(a, u)$ شباهت بین دو کاربر a و u را مشخص می کند. $\Gamma_{a,i}$ رتبه داده شده به آیت i توسط کاربر a است. $\bar{\Gamma}_a$ میانگین رتبه های کاربر a است و n تعداد کل آیت ها در فضای کاربر-آیت است.

$$TF_u = \{v | S_{u,v} > \theta_t, v \in U\} \quad (2)$$

در رابطه فوق، TF_u مجموعه ای از دوستان مورد اعتماد v کاربر فعال u است. در واقع، این افراد از میان تمامی کاربران U موجود که مقدار شباهت $S_{u,v}$ آن ها از مقدار آستانه θ_t بیشتر است انتخاب می شوند؛ بنابراین، چنین کاربرانی به عنوان دوستان مورد اعتماد کاربر فعال u در نظر گرفته خواهند شد. در روش پیشنهادی مقدار پیش فرض آستانه (θ_t) صفر در نظر گرفته شده است.

لازم به ذکر است که معیار شباهت پیرسون دو کاربر (گردشگر) یا دو کالا (هتل) را از منظری بررسی می نماید که آن ها چقدر از لحاظ خطی به یکدیگر مرتبط می باشند. از این رو ضریب همبستگی پیرسون، همبستگی هتل هایی که به طور مشترک توسط گردشگر u و v امتیاز داده شده باشند را محاسبه می نماید.

-ماتریس زمینه: این ماتریس بر اساس اطلاعات زمینه سیستم استخراج می شود. زمینه به صورت عواملی که در توصیف اطلاعات زمینه ای وجود دارند در ۵ دسته قرار می گیرند: فردی، فعالیت، روابط، زمان و مکان. در یک تعریف معمول، زمینه عبارت است از "مجموعه ای از متغیرها که ممکن است مورد علاقه یک عامل باشد و در فعالیت هایش تأثیر بگذارد." در مجموعه داده TripAdvisor که یک مجموعه داده اطلاعات هتل ها و رتبه های کاربران مختلف به آن-هاست، هر هتل به عنوان یک آیت در نظر گرفته می شود و هر کاربر به معیارهای مختلف آن هتل اعم از موقعیت مکانی، تمیزی، اتاق ها، سرویس دهی و ... به عنوان اطلاعات زمینه یک امتیاز بین ۰ تا ۵ می دهد. در جدول (۱) قسمتی از مجموعه داده نشان داده شده است. در شکل (۲) نیز الگوریتم شبه کد فاز پیش پردازش داده های روش پیشنهادی نمایش داده شده است.

جدول ۱. نمایی از اطلاعات زمینه در مجموعه داده خلاصه شده

User ID	Overall Rating	Value Aspect	Rooms Aspect	Location Aspect	Cleanliness Aspect	Hotel ID
۱۱	۵	۱-	۵	۵	۵	Hotel_572859
۱۴	۵	۵	۵	۵	۵	Hotel_566077
۱۸	۵	۵	۴	۱-	۵	Hotel_565550
۲۰	۳	۳	۴	۲	۵	Hotel_566077
۲۲	۴	۳	۳	۳	۴	Hotel_570888
۴۵	۵	۵	۵	۵	۵	Hotel_570888

Algorithm 1. The pseudo-code of Preprocessing Data of the proposed Approach

Input: Parameters TripAdvisor Dataset

Output: Rate Matrix, Similarity Matrix, Trust Matrix, Context Matrix

Algorithm:

1: *Read The dataset;*

2: *Create a Matrix with n Row and m Column;*

3: *Fill the Rate Matrix*

4: *Calculate similarity matrix for all user*

5: *Calculate Trust matrix using similarity matrix*

6: *Read The Context data*

7: *Create a Matrix with n Row and m Column;*

8: *Fill the Context Matrix*

9: *End;*

شکل ۲. شبه کد فاز پیش پردازش داده‌های روش پیشنهادی

خوشه‌بندی کاربران یا هتل‌ها

در این قسمت با استفاده از اطلاعات زمینه تلاش می‌شود که کاربران (گردشگران) و آیتم‌های (هتل‌های) سیستم به تعدادی خوشه تقسیم‌بندی شده و سپس بر اساس این خوشه‌ها به کاربران پیشنهاد گردد.

در این مرحله، یک روش خوشه‌بندی جدید و مبتنی بر تئوری گراف به منظور دسته‌بندی کاربران و یا آیتم‌های موجود در سیستم ارائه شده است. روش خوشه‌بندی ارائه شده در واقع مهم‌ترین فاز روش پیشنهادی را تشکیل می‌دهد که شامل سه مرحله است.

اولین قدم در خوشه‌بندی، ارائه گراف سیستم است. در این مرحله، مجموعه کاربران و یا آیتم‌ها به صورت یک گراف $G = (V, E, W)$ با استفاده از اطلاعات زمینه نگاشت می‌شوند که V مجموعه کاربران و یا آیتم‌ها را به عنوان رئوس گراف نشان می‌دهد، E و W به ترتیب مجموعه یال‌ها و وزن‌های شباهت بین هر جفت کاربر و یا آیتم می‌باشند.

گراف مورد نظر می‌تواند بر اساس اطلاعات کاربران و یا آیتم‌های موجود در سیستم ساخته شود. در روش پیشنهادی مبتنی بر کاربر، ترکیبی از معیار شباهت پیرسون، اطلاعات زمینه و روابط اعتماد به عنوان مقادیر شباهت نهایی بین کاربران مورد استفاده قرار می‌گیرد. مقدار وزن نهایی بین هر جفت کاربر با استفاده از رابطه (۳) قابل محاسبه است.

$$w_{u,v} = \sigma \cdot \text{trust}(u, v) + (1 - \sigma) \cdot \text{sim}(u, v) \quad (3)$$

پارامتر σ وزن مقادیر شباهت و اعتماد را کنترل می‌کند. در روش پیشنهادی بر اساس تکنیک‌های سعی و تکرار برای این پارامتر مقدار $\sigma = 0.5$ در نظر می‌گیریم تا وزن یکسانی به شباهت و اعتماد داده باشیم. $\text{trust}(u, v)$ مقدار اعتماد بین کاربر u و کاربر v است و با استفاده از رابطه (۴) محاسبه می‌نماییم (فرمال و لکرون، ۲۰۱۷):

$$t_{u,v} = \left(\frac{d_{\max} - d_{u,v} + 1}{d_{\max}} \right) \quad (4)$$

در این رابطه، $t_{u,v}$ نشان‌دهنده مقدار اعتماد بین کاربر u و v است. این مقدار اعتماد با توجه به فاصله کاربر u از v مشخص خواهد شد. d_{\max} حداکثر عمق انتشار و $d_{u,v}$ فاصله کاربر فعال u از هر یک از کاربران مورد اعتماد v است. در رابطه (۴) مقدار d_{\max} بهینه را می‌توان از رابطه (۵) به دست آورد، که در آن میانگین طول مسیرهای شبکه بوده و n تعداد گره‌های موجود در شبکه و k نشان‌دهنده میانگین درجه گره‌ها در شبکه اعتماد است (یوان^۱ و همکاران، ۲۰۱۰).

در واقع میزان اعتماد بین دو کاربر (گردشگر) بیان‌کننده این مفهوم است که گردشگران قبلی که از مکان‌های مختلف گردشگری بازدید کرده و به آن‌ها امتیاز داده‌اند تا چه اندازه می‌توانند در انتخاب مکان مناسب برای گردشگران فعلی مورد استفاده واقع شوند.

$$d_{\max} = \lceil L^R \rceil = \left\lceil \frac{\ln(n)}{\ln(k)} \right\rceil \quad (5)$$

و در نهایت $\text{sim}(u, v)$ مقدار شباهت بین کاربر u و کاربر v را نشان می‌دهد که با استفاده از اطلاعات زمینه و از رابطه (۶) قابل محاسبه است.

$$\begin{aligned} \text{sim}(u, v) &= (1 - \gamma) \frac{\sum_{i \in A_{u,v}} (r_i(u) - \bar{r}(u))(r_i(v) - \bar{r}(v))}{\sqrt{\sum_{i \in A_{u,v}} (r_i(u) - \bar{r}(u))^2} \sqrt{\sum_{i \in A_{u,v}} (r_i(v) - \bar{r}(v))^2}} \\ &+ \gamma \frac{1}{(G_u - G_v) + (M_u - M_v)} \end{aligned} \quad (6)$$

که $r_i(u)$ رتبه داده شده به آیت i توسط کاربر u است، $\bar{r}(u)$ میانگین رتبه‌های داده شده توسط کاربر u است و $A_{u,v}$ مجموعه آیت‌هایی است که توسط هر دو کاربر u و v رتبه داده

شده‌اند. دو متغیر G و M اطلاعات مربوط به زمینه افراد است که موقعیت و وضعیت روحی فرد را نشان می‌دهد.

در رابطه (۶) γ یک پارامتر ثابت است که میزان تأثیر اطلاعات زمینه را نشان می‌دهد. اگر میزان پارامتر γ برابر ۱ باشد در محاسبه شباهت فقط اطلاعات زمینه افراد لحاظ می‌گردد و اگر مقدار پارامتر γ برابر صفر در نظر گرفته شود اطلاعات زمینه در نظر گرفته نمی‌شود.

در اینجا اطلاعات زمینه‌ای گردشگر که همان موقعیت مکانی‌اش است مدنظر بوده که توسط این پارامتر بر میزان شباهت بین دو گردشگر تأثیر خود را نشان می‌دهد.

در روش پیشنهادی نیز از ترکیب معیار شباهت پیرسون و اطلاعات زمینه برای محاسبه مقادیر شباهت بین کاربران استفاده می‌شود. به‌عنوان مثال در مجموعه داده‌ای مورد استفاده، هر هتل به‌عنوان یک آیتم در نظر گرفته می‌شود که هر کاربر به معیارهای مختلف آن هتل اعم از موقعیت مکانی^۱، تمیزی^۲، اتاق‌ها^۳، سرویس‌دهی^۴ و ... یک امتیاز بین ۰ تا ۵ می‌دهد، بنابراین مقادیر شباهت بین کاربر u و کاربر v را با استفاده از رابطه (۷) می‌توان به دست آورد.

$$\begin{aligned} \text{sim}(u, v) & \quad (7) \\ &= (1 - \gamma) \frac{\sum_{i \in A_{u,v}} (r_i(u) - \bar{r}(u))(r_i(v) - \bar{r}(v))}{\sqrt{\sum_{i \in A_{u,v}} (r_i(u) - \bar{r}(u))^2} \sqrt{\sum_{i \in A_{u,v}} (r_i(v) - \bar{r}(v))^2}} \\ &+ \gamma \frac{1}{(L_i - L_j) + (C_i - C_j) + (R_i - R_j) + (S_i - S_j)} \end{aligned}$$

[[

در اینجا پارامترهای زمینه شامل L_i جهت موقعیت مکانی هتل، C_i تمیزی، R_i اتاق‌های هتل و S_i سرویس‌دهی هتل می‌باشند.

1. Location

2. Cleaness

3. Rooms

4. Services

در روش پیشنهادی مبتنی بر آیتم، فقط از معیار شباهت پیرسون برای محاسبه مقادیر شباهت بین آیتم‌ها استفاده می‌شود. تابع شباهت پیرسون مقادیر شباهت بین آیتم i و آیتم j را با استفاده از رابطه (۸) به دست می‌آورد.

$$\text{sim}(i, j) = (1 - \gamma) \frac{\sum_{u \in U_i \cap U_j} (r_i(u) - \bar{r}_i)(r_j(u) - \bar{r}_j)}{\sqrt{\sum_{v \in U_i} (r_i(v) - \bar{r}_i)^2} \sqrt{\sum_{w \in U_j} (r_j(w) - \bar{r}_j)^2}} \quad (8)$$

که $r_i(u)$ ، رتبه داده شده به آیتم i توسط کاربر u است، U_i مجموعه کاربرانی است که به آیتم i رتبه داده‌اند، و \bar{r}_i میانگین رتبه‌های داده شده به آیتم i است که با استفاده از رابطه زیر قابل محاسبه است:

m ، تعداد کل کاربران موجود در سیستم است.

$$\bar{r}_i = \frac{1}{m} \cdot \sum_{l=1}^m r_{l,i} \quad (9)$$

در روش پیشنهادی از الگوریتم مبتنی بر محدودیت معرفی شده در (بهمنی و همکاران، ۲۰۱۲) برای یافتن چگالی‌ترین زیرگراف استفاده شده است. از زیرگراف به دست آمده به عنوان مجموعه مراکز اولیه الگوریتم خوشه‌بندی ارائه شده استفاده می‌شود. هدف اصلی این مرحله، یافتن یک زیرمجموعه از رئوس $\tilde{S} \subseteq V$ با حداقل تعداد رئوس k است که میانگین چگالی آن‌ها مینیمم باشد. در روش پیشنهادی، الگوریتم (بهمنی و همکاران، ۲۰۱۲) برای یافتن چگالی‌ترین زیرگراف، به گونه‌ای تغییر داده شده که یک زیرگراف با کمترین چگالی به دست آورد.

شبه کد الگوریتم تغییر یافته مبتنی بر محدودیت برای یافتن چگالی‌ترین زیرگراف در شکل (۳) آورده شده است. که بر روی گراف ورودی G اجرا شده و چگالی آن را محاسبه می‌کند.

رئوسی که درجه آن‌ها بیشتر از یک حد آستانه θ باشد، به عنوان رئوس کاندید تشخیص داده می‌شوند که می‌توانند از گراف ورودی حذف شوند.

Algorithm2 . The modified dense subgraph finding

Require : $G = (V, E), k > 0, \text{ and } r > 0$

Algorithm:

1: $S, \tilde{S} \leftarrow V;$

2: while $S \neq \emptyset$ do

3: $\tilde{A}(S) \leftarrow \{i \in S \mid \text{deg}_S(i) \geq 2 * \rho(S)\};$

4: Let $A(S) \subseteq \tilde{A}(S)$, with $|A(S)| = r * |\tilde{A}(S)|$;

5: $S \leftarrow S \setminus A(S);$

6: if $|S| \geq k$ and $\rho(S) < \rho(\tilde{S})$ then

7: $\tilde{S} \leftarrow S;$

8: end if

9: end while

10: return $\tilde{S};$

شکل ۳. شبه کد محاسبه چگالی زیرگراف (بهمینی و همکاران، ۲۰۱۲)

رئوس کاندید با $\tilde{A}(S)$ نشان داده می‌شوند و مقدار آستانه به صورت $\theta = 2 * \rho(S)$ تعریف می‌گردد. سپس تعدادی از رئوس موجود در مجموعه کاندید که با $A(S)$ نشان داده می‌شوند، از مجموعه کاندید حذف می‌شوند. در صورتی که زیرگراف به دست آمده غیر تهی باشد، الگوریتم بر روی گراف باقیمانده اجرا می‌شود. در نهایت، مجموعه رئوس به مجموعه $S \setminus A(S)$ کاهش می‌یابد و الگوریتم تضمین می‌کند که زیرگراف به دست آمده حداقل دارای k رأس است.

مراکز اولیه خوشه‌ها که در مرحله قبل به دست آمده‌اند، ممکن است مراکز بهینه نهایی نباشند. برای رفع این مشکل، یک روش جدید برای یافتن مراکز خوشه بهتر ارائه شده است. بدین منظور، ابتدا مجموع مقادیر شباهت بین یک مرکز خوشه کاندید $v_i \in C_j$ و دیگر اعضای موجود در خوشه، $(\text{sum}(v_i) = \sum_{v_t \in C_j, v_t \neq v_i} \text{sim}(v_i, v_t))$ محاسبه می‌شود. این مقدار برای تمامی اعضای موجود در خوشه‌ها محاسبه می‌گردد. در نهایت، برای هر خوشه، یک رأس با بیشترین مقدار محاسبه‌شده به عنوان مرکز خوشه جدید بر اساس رابطه (۱۰) انتخاب می‌شود.

$$\text{newcenter} = \underset{v_i \in C_j}{\text{argmax}} \text{sum}(v_i) \quad (10)$$

پس از انتخاب مراکز خوشه جدید برای تمامی خوشه‌ها، سایر رئوس گراف به نزدیک‌ترین خوشه تعلق می‌گیرند. این فرایند تا زمانی که هیچ کدام از مراکز خوشه‌ها تغییر نکنند، به صورت تکراری اجرا می‌گردد. در پایان، پس از به دست آمدن مراکز خوشه جدید، تمامی رئوس که به عنوان مرکز خوشه انتخاب نشده‌اند، به نزدیک‌ترین مرکز خوشه موجود تعلق خواهند گرفت. در واقع با استفاده از تکنیک خوشه‌بندی ارائه شده در این مقاله می‌توان گردشگران را بر اساس سلايق مشترکشان به خوشه‌های مختلفی تقسیم کرده و از این طریق امکان محاسبه دقیق‌تر شباهت بین گردشگر هدف و گردشگران مشابهش را فراهم کرد.

در پایان مرحله دوم فاز خوشه‌بندی، برخی از خوشه‌های تشکیل شده ممکن است دارای تعداد کافی اعضا برای استفاده در مرحله پیش‌بینی رتبه نباشند. این نکته را باید در نظر داشت که خوشه‌های تشکیل شده با تعداد کم اعضای موجود در آن‌ها، باعث می‌شوند که برای یک کاربر هدف موجود در این خوشه‌ها، تعداد کمی کاربر به عنوان کاربران همسایه در نظر گرفته شود؛ بنابراین، تعداد کم همسایه‌های موجود برای کاربر هدف علاوه بر اینکه باعث ایجاد پیش‌بینی‌های با دقت پایین برای این کاربر می‌شود، باعث کاهش نرخ پوشش رتبه‌ها نیز می‌گردد.

برای غلبه بر این مشکل، در مرحله سوم فاز خوشه‌بندی، یعنی مرحله ادغام خوشه‌ها، خوشه‌هایی که دارای تعداد اعضای کمتر از یک حد آستانه باشند با سایر خوشه‌های موجود ادغام

می شوند. اعضای موجود در خوشه های ادغام شده، به نزدیک ترین خوشه بعدی تعلق خواهند گرفت و خوشه قبلی حذف می گردد.

پیش بینی رتبه

در این بخش، یک مجموعه از هتل های موردعلاقه گردشگر هدف که رتبه آن ها بر اساس خوشه های به دست آمده از فاز خوشه بندی پیش بینی شده اند، به وی پیشنهاد می شود. در مرحله پیش بینی رتبه برای روش پیشنهادی مبتنی بر آیتام، رتبه داده شده به هتل i توسط گردشگر u با استفاده از رابطه (۱۱) قابل محاسبه است.

$$p_i(u) = \bar{r}_i + \frac{\sum_{j \in C_i} \text{sim}(i, j) \cdot (r_j(u) - \bar{r}_j)}{\sum_{j \in C_i} |\text{sim}(i, j)|} \quad (11)$$

که $r_j(u)$ ، رتبه داده شده به هتل j توسط گردشگر u است، \bar{r}_i میانگین رتبه های داده شده به هتل i است، $\text{sim}(i, j)$ یک تابع شباهت بین هتل i و هتل j است که با استفاده از رابطه (۸) قابل محاسبه است، و C_i مجموعه هتل های موجود در خوشه مربوط به هتل i است.

در روش پیشنهادی مبتنی بر کاربر، رتبه داده شده به هتل i توسط گردشگر u با استفاده از رابطه (۱۲) قابل محاسبه است.

$$p_i(u) = \bar{r}(u) + \frac{\sum_{v \in C_u} w_{u,v} \cdot (r_i(v) - \bar{r}(v))}{\sum_{v \in C_u} |w_{u,v}|} \quad (12)$$

که $r_i(v)$ ، رتبه داده شده به هتل i توسط گردشگر v است، $\bar{r}(u)$ میانگین رتبه های داده شده توسط گردشگر u است، $w_{u,v}$ تابع شباهت بین گردشگر u و گردشگر v است (رابطه ۳)، و C_u مجموعه گردشگران موجود در خوشه مربوط به گردشگر u را نشان می دهد. در نهایت، بر اساس رتبه های پیش بینی شده، N هتل با بیشترین رتبه (Top-N) به عنوان هتل های موردعلاقه گردشگر هدف به وی پیشنهاد می شوند.

شبه کد فاز دوم و سوم روش پیشنهادی در شکل (۴) نشان داده شده است. ورودی‌های الگوریتم پیشنهادی شامل پارامترهای k ، r ، m و top_N می‌باشند که به ترتیب، مینیمم اندازه مجموعه رتوس کاهش یافته، نسبت تعداد گردشگران و یا هتل‌هایی که در هر تکرار از الگوریتم باید حذف شوند، مقدار آستانه برای اعضای خوشه‌ها در مرحله ادغام خوشه‌ها و اندازه لیست توصیه را نشان می‌دهند.

Algorithm 3. The pseudo-code of the proposed methods

Input: Parameters k , r , m , and top_N .

Output: Top-N recommendation list.

Algorithm:

- ۱: Split dataset into train set Tr and test set Te ;
 - ۲: Map all users/items to a graph $G = (V, E, W)$ using the train set Tr ;
 - ۳: Apply modified dense subgraph finding algorithm with constraint k i.e. Algorithm 1) to obtain \tilde{S} as initial center set;
 - ۴: $k' = |\tilde{S}|$;
 - ۵: Set $p_j = \tilde{S}_j, \forall j = 1, \dots, k'$;
 - ۶: Let $p_j, \forall j = 1, \dots, k'$ be initial center corresponding to j -th cluster C_j ;
 - ۷: Associate each non-selected user/item to nearest cluster;
 - ۸: Select new centers $p'_j = \arg \max_{v_i \in C_j} \text{sum}(v_i), j = 1, \dots, k'$, where

$$\text{sum}(v_i) = \sum_{v_t \in C_j, v_t \neq v_i} \text{sim}(v_i, v_t);$$
 - ۹: if $p_j = p'_j, \forall j = 1, \dots, k'$ then goto line 10, else $p_j = p'_j, \forall j = 1, \dots, k'$ and goto line 7;
 - ۱۰: for all $C_j, j = 1, \dots, k'$ do
-

```

:۱۱ if  $|C_j| < m$  then
    :۱۲ Merge  $C_j$  to other clusters;
:۱۳ end if
:۱۴ end for
:۱۵ for all  $r_i(u) \in Te$  do
    :۱۶ Predict the rate  $p_i(u)$  of the item  $i$  for the user  $u$ ;
:۱۷ end for
:۱۸ Recommend  $top\_N$  of items as the recommendation list to the user  $u$ ;

```

شکل ۴. شبه کد فاز دوم و سوم روش پیشنهادی

تجزیه و تحلیل نتایج

سایت TripAdvisor از بزرگ‌ترین شبکه‌های اجتماعی گردشگران در اینترنت است که در آن بیش از ۵۷۰ میلیون نظر^۱ و راهنما درباره ۷,۳ میلیون مقصد اقامتی، گردشگری، رستوران‌ها، کافه‌ها ارائه شده است. ماهانه نیم میلیارد کاربر اینترنت از این سایت و مجموعه سایت‌های آن بازدید می‌کنند. جدول (۲) نمایی از این مجموعه داده را نشان می‌دهد.

جدول ۲. مجموعه داده TripAdvisor

File	Features
Hotels	Name, hotel URL, price, hotel ID, img URL
Users and reviews	authorLocation,title,author,reviewID,reviewText,date,overall,value , rooms, location, cleanness, service, sleep Quality

این مجموعه داده شامل اطلاعات هتل‌ها و رتبه‌های کاربران مختلف به این هتل‌ها است. در این مجموعه داده‌ای هر هتل به عنوان یک آیتم در نظر گرفته می‌شود که هر کاربر به معیارهای

مختلف آن هتل اعم از موقعیت مکانی، تمیزی، اتاق‌ها، سرویس ده و ... یک امتیاز بین ۰ تا ۵ می‌دهد و البته رتبه ۱- نیز نشان‌دهنده گم‌شدن رتبه در فایل html اصلی است. به منظور استفاده از اطلاعات مجموعه داده قطعه کدی به زبان #C نوشته شد و اطلاعات هتل‌ها و نظرات کاربران به XML تبدیل گردید، سپس از مجموعه داده جدید خلاصه شده به منظور تولید ماتریس رتبه استفاده گردید.

معیارهای متفاوتی را می‌توان به منظور ارزیابی سیستم‌های توصیه‌گر ارائه نمود که در تمامی آن‌ها مهم‌ترین معیارها بررسی میزان خطا و پوشش هستند. از جمله این معیارها، می‌توان به معیار میانگین خطای مطلق^۱ و میانگین خطای مطلق کاربر^۲ برای اندازه‌گیری دقت پیش‌بینی رتبه‌ها اشاره نمود که در ادامه روابط مربوط به هر یک را نشان خواهیم داد.

$$MAE = \frac{\sum_u \sum_i |r_{u,i} - p_{u,i}|}{N} \quad (13)$$

در رابطه فوق، N نشان‌دهنده تعداد رتبه‌های موجود در ماتریس کاربر-آیتم، $r_{u,i}$ مقدار رتبه واقعی کاربر u به آیتم i و $p_{u,i}$ مقدار رتبه پیش‌بینی شده برای آیتم i توسط سیستم توصیه‌گر است.

$$MAUE = \frac{\sum_{u \in U} MAE_u}{N} \quad (14)$$

در این رابطه N نشان‌دهنده تعداد کاربران و MAE_u میانگین خطای مطلق برای کاربر u است.

در معیار پوشش، میزان پوشش کاربران^۳ و آیتم‌ها^۴ با توجه به رتبه‌بندی انجام شده مورد بررسی قرار می‌گیرد. نتیجه این معیارها، مقداری بین ۰ و ۱ بوده و هر چه مقدار به ۱ نزدیک‌تر باشد نشان‌دهنده پوشش مناسب برای هر یک از معیارهای پوشش کاربران و آیتم‌ها خواهد بود. در

1. Mean Absolute Error (MAE)

2. Mean Absolute User Error) MAUE)

3. User Coverage (UC)

4. Rate Coverage (RC)

ادامه روابط این معیارها را ارائه می‌کنیم.

$$UC = \frac{N_i}{N_t} \quad (15)$$

در این رابطه، UC مقدار پوشش کاربران، N_i نشان‌دهنده تعداد کل کاربران فعالی است که سیستم توصیه گر توانسته حداقل یک رتبه را برای آن‌ها پیش‌بینی کند و N_t تعداد کل کاربران فعال را نشان می‌دهد.

$$RC = \frac{N_r}{N_c} \quad (16)$$

در رابطه فوق، RC مقدار پوشش آیتم‌ها، N_r نشان‌دهنده تعداد کل آیتم‌هایی است که سیستم توصیه گر توانسته مقداری به‌عنوان رتبه را برای آن‌ها پیش‌بینی کند و N_c تعداد کل آیتم‌هایی است که قصد پیش‌بینی رتبه را برای آن‌ها داشته‌ایم.

در روش پیشنهادی از اطلاعات زمینه، اعتماد و خوشه‌بندی زیرگراف متراکم جهت کاهش خطا استفاده شده است. نتایج حاصل از روش پیشنهادی با توجه به تعداد ۵ خوشه به دست آمده از الگوریتم زیرگراف متراکم ارائه می‌گردد. در این بخش عملکرد روش پیشنهادی با روش‌های پالایش گروه (CF)، مبتنی بر اعتماد (MT) (ماسا و آوسانی^۱، ۲۰۰۷)، مقاله (آزادجلال و همکاران، ۲۰۱۷) به‌عنوان مقاله پایه و مقاله‌های (ژانگ و موریموتو^۲، ۲۰۱۷؛ عبادی و کرزیزاک^۳، ۲۰۱۶) مقایسه شده است.

با توجه به نتایج به دست آمده برای روش‌ها و معیارهای متفاوت، روش CF در مقایسه با سایر روش‌ها، دارای مقدار پوشش پایینی برای کاربران و آیتم‌ها در کاربران شروع سرد است که این موضوع به دلیل استفاده از معیار شباهت در روش مورد نظر است. چراکه به دلیل شباهت کم این نوع از کاربران با سایر کاربران موجود، مقدار شباهت معمولاً صفر شده و منجر به عدم پیش‌بینی رتبه‌ها برای آیتم‌ها و کاربران خواهد شد. در نتیجه میزان پوشش برای کاربران شروع سرد کاهش خواهد یافت.

1. Massa & Avesani

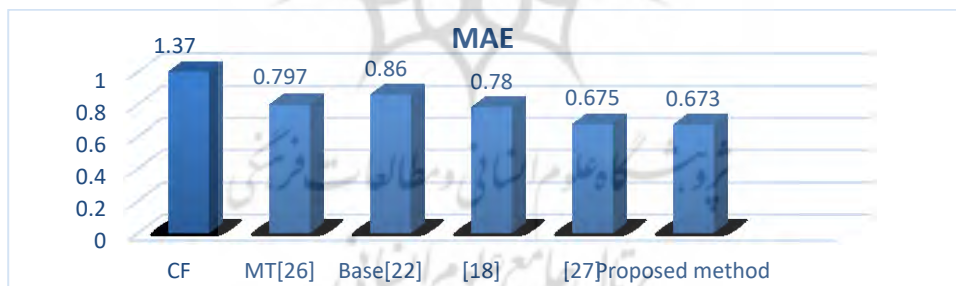
2. Zhang & Morimoto

3. Krzyzak

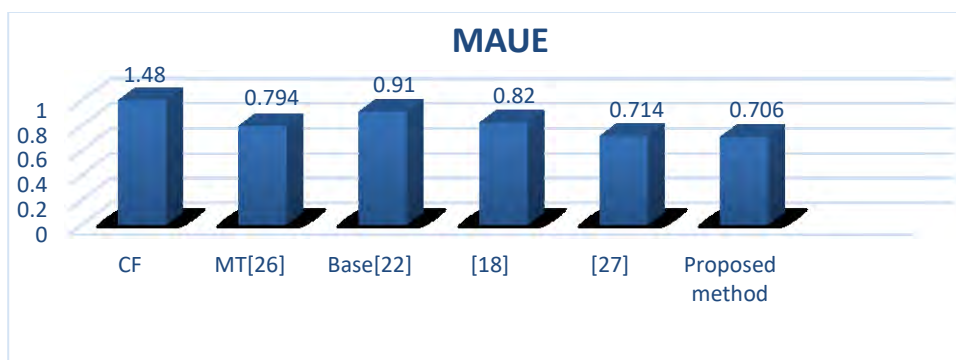
در روش MT و روش پیشنهادی به دلیل استفاده از معیار اعتماد و انتشار اعتمادی که در شبکه اعتماد دارند مشکل موجود را رفع کرده‌اند. علاوه بر این، می‌بینیم که روش پیشنهادی در مقایسه با سایر روش‌ها از MAE و MAUE کمتری برخوردار است. این مسئله نشان‌دهنده این موضوع است که روش ارائه شده دارای خطای کمتری در پیش‌بینی رتبه‌ها بوده و توانسته رتبه‌ها را با دقت بیشتری نسبت به سایر روش‌ها پیش‌بینی کند.

روش ارائه شده، در کنار افزایش دقت برای سیستم‌های توصیه‌گر مبتنی بر اعتماد، مقدار RC و UC را که به ترتیب نشان‌دهنده پوشش کاربران و آیتم‌ها است، در سیستم‌های توصیه‌گر مبتنی بر اعتماد حفظ کرده است. همچنین، روش پیشنهادی در کنار حفظ مقدار پوشش روش MT، میزان دقت را در پیش‌بینی رتبه‌ها نسبت به این روش بهبود داده است.

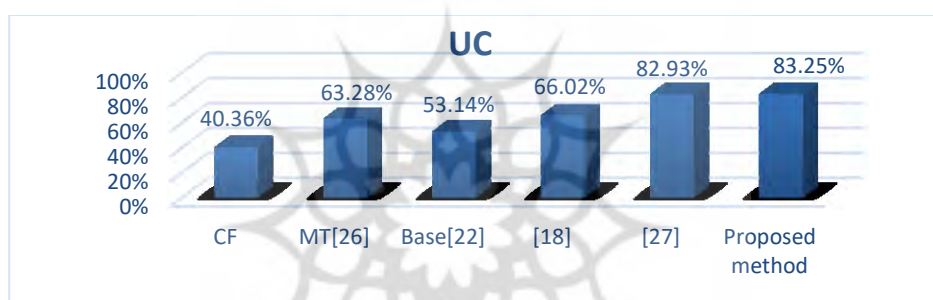
به منظور نمایش بصری نتایج به دست آمده، میزان خطا و پوشش آیتم‌ها و کاربران بر روی مجموعه داده TripAdvisor، برای تمامی کاربران و کاربران شروع سرد نشان داده شده است. بنابراین، برای تمامی کاربران در این مجموعه داده، نتایج به دست آمده برای معیارهای MAE و MAUE در شکل ۵ و ۶، نتایج به دست آمده برای معیارهای RC و UC در شکل ۷ و ۸ نشان داده شده است.



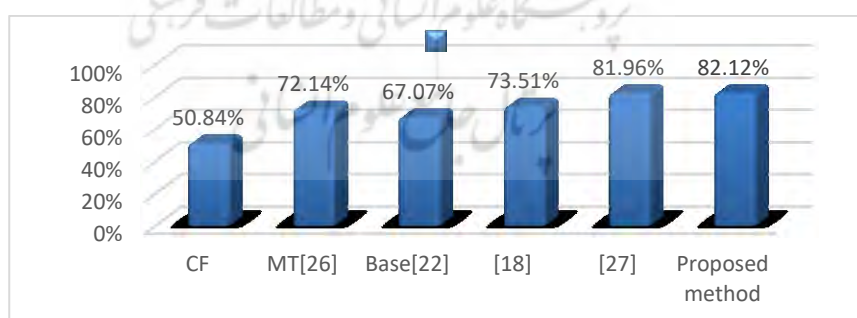
شکل ۵. مقایسه روش پیشنهادی با سایر روش‌ها بر حسب معیارهای MAE برای تمامی کاربران



شکل ۶. مقایسه روش پیشنهادی با سایر روش‌ها بر حسب معیارهای MAUE برای تمامی کاربران



شکل ۷. مقایسه روش پیشنهادی با سایر روش‌ها بر حسب معیار UC



شکل ۸. مقایسه روش پیشنهادی با سایر روش‌ها بر حسب معیار RC

نتیجه‌گیری و پیشنهادها

گردشگری الکترونیک یکی از فعالیت‌هایی است که امروزه با استفاده از شبکه گسترده اینترنت توسعه یافته است. وقتی مقصدی برای سفر انتخاب می‌شود، می‌توان سایت‌های زیادی را برای دریافت اطلاعات در مورد مکان‌های دیدنی شهر، رستوران‌ها و هتل‌ها مشاهده کرد، از طرفی انتخاب محل اقامتی متناسب با علایق گردشگر یکی از ملزومات سفر است. هرچقدر سیستم‌های توصیه‌گر بتوانند توصیه‌های بهتر و مطابق با میل و علاقه کاربران را ارائه دهند، استقبال کاربران از این گونه سیستم‌ها نیز افزایش پیدا خواهد کرد.

سیستم‌های توصیه‌گر گردشگری می‌تواند به‌طور خودکار بسیاری از فرآیندهای که نیاز به دخالت انسان دارد را به شکل بسیار مؤثری جایگزین نماید. به‌طور مثال سیستم طراحی شده در این مقاله می‌تواند به‌جای استفاده از آژانس‌های مسافرتی یا راهنمای سفر، مکان‌های مناسب اقامتی را در شهر مقصد به گردشگر معرفی نماید به‌گونه‌ای که بیشترین سنخیت را با علایق و دل‌بستگی‌های گردشگر داشته باشد.

در این مقاله با ترکیب اطلاعات زمینه، اعتماد بین کاربران و بسط آن و درنهایت خوشه‌بندی کاربران و جمع‌آوری رتبه‌های موردنیاز از کاربران مورد اعتماد، مقدار رتبه هر یک از هتل‌ها به دست آمده و بر اساس رتبه‌های پیش‌بینی شده برای کاربر هدف، N آیتم (هتل) مرتب شده بر اساس رتبه بالاتر به‌عنوان هتل‌های موردعلاقه به وی پیشنهاد می‌شود. ارزیابی روش پیشنهادی و مقایسه عملکرد آن با سایر روش‌های مورد آزمایش بر روی مجموعه داده TripAdvisor بیانگر آن است که روش پیشنهادی نتایج بهتری نسبت به سایر روش‌های مورد مقایسه دارد.

مدل ارائه شده در این مقاله به‌گونه‌ای طراحی شده است که می‌تواند از مجموعه داده‌های مختلف به‌غیر از TripAdvisor نظیر Lonelyplanet یا Expedia نیز استفاده کند بدین منظور کافی است که فاز یک فرآیند اجرایی روش پیشنهادی که برای پیش‌پردازش داده‌ها است، متناسب با مجموعه داده جدید تغییر کند.

جهت کارهای آتی در راستای این تحقیق می‌توان به استفاده از روابط فازی به منظور خوشه‌بندی کاربران و یا آیتم‌های موجود در سیستم اشاره کرد. استفاده از تکنیک‌های تشخیص جوامع مبتنی بر شبکه‌های اجتماعی یا استفاده از روابط عدم اعتماد در کنار روابط

اعتماد نیز می‌تواند از جمله پژوهش‌های آتی در خوشه‌بندی کاربران باشد. استفاده از سایر تئوری‌های گراف به منظور یافتن کاربران و یا آیت‌های مهم در سیستم همچنین می‌تواند باعث ایجاد کاربردهای دیگری در سیستم‌های توصیه گر نظیر Tripadvisor یا Booking به دست می‌آید می‌تواند مکمل یکدیگر باشد و شکاف‌های اطلاعاتی را موجود تکمیل کند. بدین منظور استفاده ترکیبی از این مجموعه اطلاعات و ارائه یک سیستم جامع توصیه گردشگری می‌تواند یکی دیگر از پتانسیل‌های تحقیقات آینده باشد. همچنین رویکرد ارائه شده در این مقاله برای پیشنهاد مکان‌های مناسب در مقاصد گردشگری می‌تواند به عنوان نرم‌افزاری کاربردی که بر روی موبایل و یا بر بستر وب قابل اجرا باشد در اختیار گردشگران قرار گیرد.



منابع

- ضیایی، محمود و فرمانی، لیلا. (۱۳۸۹). ارزیابی گردشگران اروپایی از عملکرد راهنمایان تور در ایران، فصلنامه مطالعات گردشگری، ۱۵، ۲۹-۵۲.
- فرزین، محمدرضا و نادعلی پور، زهرا. (۱۳۹۰). عوامل مؤثر بر مزیت رقابتی مقصدهای گردشگری در ایران (مورد مطالعه: منطقه چابهار). فصلنامه مطالعات گردشگری، ۱۴، ۴۱-۶۷.
- منصوری موید، فرشته و سلیمانی، سمیرا. (۱۳۹۱). ابزارهای بازاریابی و تصویر ذهنی گردشگران از مقصد، فصلنامه مطالعات گردشگری، ۱۸، ۹۳-۱۱۰.
- کاظمی، مهدی؛ نیک‌نفس، علی‌اکبر و احمد عمویی، مهدیه. (۱۳۸۴). طراحی سیستم تصمیم‌یار برای بخش اسکان صنعت گردشگری. فصلنامه مطالعات مدیریت گردشگری، ۳(۹)، ۹۱-۱۱۶.
- سهرابی، بابک؛ فضلی، صفر؛ طهماسبی پور، کاوه و رئیسی وانانی، ایمان. (۱۳۹۲). تصمیم‌گیری به‌وسیله سیستم خبره فازی برای انتخاب محل اقامت مطلوب گردشگران. مهندسی صنایع، ۴۷(۲)، ۲۰۱-۲۱۳.
- بهاری سجهرود، زهرا و طالعی، محمد. (۱۳۹۴). توسعه یک سامانه برنامه‌ریزی گردشگری بر پایه مدیریت زمانی و مکانی. فصلنامه علمی- پژوهشی اطلاعات جغرافیایی «سپهر»، ۲۴(۹۳)، ۴۲-۵۲.
- ابوالحسینی، سینا و آل‌شیخ، علی‌اصغر. (۱۳۹۷). توسعه یک سامانه همراه توصیه‌گر و برنامه‌ریزی مسیر برای گردشگرهای انفرادی. نشریه علمی پژوهشی علوم و فنون نقشه‌برداری، ۷(۴)، ۸۹-۱۰۱.
- World Tourism Organization (UNWTO) and World Travel and Tourism Council (WTTC). (2017). "The Travel & Tourism Competitiveness Report 2017", http://www3.weforum.org/docs/WEF_TTCR_2017_web_0401.pdf. [Accessed 15/04/2019].
- L. Kzaz, D. Dakhchoune, D. Dahab, (2018), "Tourism recommender systems: an overview of recommendation approaches", International Journal of Computer Applications, 180(20), 9-13.
- Ricci, F., Rokach L., Shapira, B. and Kantor, P. B. (2011). *Recommender Systems Handbook*. Springer.
- Jinming, H. (2010). Application and research of collaborative filtering in e-commerce recommendation system. *Computer Science and Information Technology (ICCSIT)*, 3rd IEEE International Conference, 686-689.
- Shepitsen, A. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. *in Proceedings of the ACM conference on Recommender systems RecSys*, 259-266.

Schein, A. I. (2002). Methods and metrics for cold-start recommendation. *in Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 253-260.

Schilit, B. Adams, N. and Want, R. (1994). Context-aware computing applications. *Mobile Computing Systems and Applications. Proceedings of the First Workshop on Mobile Computing Systems and Applications*, 85 – 90.

Kelleher, J. and Bridge, D. (2004). An accurate, scalable collaborative recommender. *Artificial Intelligence Review*, 21,193-213.

Lucas, J. (2013). A hybrid recommendation approach for a tourism system. *Expert Systems with Applications*, 40, 3532-3550.

Uchytig, G. and Clark, K. (2004). Hierarchical agglomerative clustering for agent-based dynamic collaborative filtering. *Intelligent Data Engineering and Automated Learning – IDEAL*, 827-833.

Wang, J. Zhang, N. Y. & Yin, J. (2010). Collaborative filtering recommendation based on fuzzy clustering of user preferences. *in Seventh international conference on fuzzy systems and knowledge discovery (FSKD)*, 1946–1950.

Haruna, K., Ismail, M., Suhendroyono, A. S., Damiasih, D., Pierewan, A., Chiroma, Herawan, C. H. T. (2017). Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction. *Applied Sciences*, 7(12),1211-1221.

Viktoratos, I., Tsadiras, A., & Bassiliades, N. (2018). Combining Community-Based Knowledge with Association Rule Mining to Alleviate the Cold Start Problem in Context-Aware Recommender Systems. *Expert Systems With Applications*, 101, 78-90.

Hernando, A. Bobadilla, J. Ortega, F., & Tejedor, J. (2013), "Incorporating reliability measurements into the predictions of a recommender system. *Information Sciences.*, 218, 1-16.

Puntheeranurak, S. & Tsuji, H. (2007). A multi-clustering hybrid recommender system. *in Proceedings of the 7th IEEE International Conference on Computer and Information Technology*.

Leal, F. (2017). Trust-based Modelling of Multi-criteria Crowdsourced Data. *Data Science and Engineering*, 2,199-209.

Gavalas, D., Kasapakis, V., Konstantopoulos, C., Pantziou, G., Vathis, N., & Zaroliagis, C. (2015). The eCOMPASS multimodal tourist tour planner. *Expert Systems with Applications*, 42, 7303-7316.

Tiwari, S., & Kaushik, S. (2014). Information Enrichment for Tourist Spot Recommender System Using Location Aware Crowdsourcing. *IEEE 15th International Conference on Mobile Data Management*, 2, 11-14.

Huang, H. (2016). Context-Aware Location Recommendation Using Geotagged Photos in Social Media. *ISPRS International Journal of Geo-Information*, 5(11), 195-206.

Azadjalal, M.M. (2017). A trust-aware recommendation method based on Pareto dominance and confidence concepts. *Knowledge-Based Systems*, 116, 130-139.

Mohammadnezhad, M. & Mahdavi, M. (2012). An effective model for improving the quality of recommender systems in mobile e-tourism. *International Journal of Computer Science & Information Technology (IJCSIT)*, 4(1), 83-92.

Zhenxing, X., Chen, L. and Chen, G. (2015). Topic based context-aware travel recommendation method exploiting geotagged photos. *Neurocomputing*, 155, 99-107.

Frémal, S. and Lecron, F. (2017). Weighting strategies for a recommender system using item clustering based on genres. *Expert Systems with Applications*, 77, 105-113.

Yuan, W. (2010). Improved trust-aware recommender system using small-worldness of trust networks. *Knowledge-Based Systems*, 23, 232-238.

Bahmani, B. Kumar, R. and Vassilvitskii, S. (2012). Densest subgraph in streaming and mapreduce. in *Proceedings of VLDB Endowment*, 5(5), 454-465.

Massa, P., & Avesani, P. (2007). Trust-aware recommender systems. *The Proceedings of the ACM conference on Recommender systems*, 17-24.

Zhang, Z. & Morimoto, Y. (2017). Collaborative Hotel Recommendation based on Topic and Sentiment of Review Comments. *Proceedings of the DEIM*, 2-6.

Ebadi, A. & Krzyzak, A. (2016). A Hybrid Multi-Criteria Hotel Recommender System Using Explicit and Implicit Feedbacks. *Proceedings of the 18th International Conference on Applied Science in Information Systems and Technology (ICASIST)*, 10(8), 1377-1385.